Proskauer Rose LLP | 1001 Pennsylvania Avenue, NW | Washington, DC 20004

Colin R. Kass
202-416-6890
ckass@proskauer.com

October 30, 2024

Via ECF

Hon. William H. Alsup
United States District Court for the Northern District of California
Courtroom 12, 19th Floor
450 Golden Gate Ave
San Francisco, CA 94102

Re: *X Corp. v. Bright Data, 23-cv-03698 (N.D. Cal.)*
Bright Data's Letter Regarding Its Anticipated Motion to Dismiss X's New Claims

Dear Judge Alsup:

As requested at the October 23, 2024 hearing, Bright Data writes to preview the grounds for dismissing the three new claims in X's proposed Second Amended Complaint (ECF 118-1): the Computer Fraud and Abuse Act (CFAA), the California Comprehensive Computer Data Access and Fraud Act (CDAFA), and the Digital Millenium Copyright Act (DMCA) claims.

A. X Plausibly Alleges Access to Public Data Only.

X's new claims fail because they are brought under anti-hacking statutes,[1] and accessing public data is not hacking. As the Ninth Circuit held, "when a computer network generally permits public access to its data, a user's accessing that publicly available data will not constitute access without authorization under the CFAA." *hiQ Labs, Inc. v. LinkedIn Corp.*, 31 F.4th 1180, 1201 (9th Cir. 2022); *Hattler v. Ashton*, 2017 WL 11634742, *8 (C.D. Cal. 2017) (DMCA claim fails where "the Works are accessible to the public on many websites"). X's new claims ignore this and seek to punish Bright Data for accessing public information X itself has chosen to "disclose … as broadly as possible." X Privacy Policy § 3.1.

X does not allege that Bright Data engages in hacking. To the contrary, the SAC plausibly alleges only access to and copying of information publicly available on X's site. "[S]pecific post[s]," their "number of likes, replies and reposts," the "profile of the person who posted the content," "a curated list of other posts from that individual," other "linked posts," and Google search results are all publicly available without a login. SAC ¶¶ 64-66. X does not identify any information beyond this that Bright Data scraped. ECF 120-2 at 12-13 (explaining that X "fails to identify ***any*** information that Bright Data scrapes that is behind a log-in"). Nor is X's alleged skepticism about the scope of Bright Data's scraping a well-pled allegation of logged-in activity.

[1] *In re BetterHelp, Inc. Data Disclosure Cases*, 2024 WL 3416511, *5 (N.D. Cal. 2024) ("Both the CDAFA and the CFAA were 'enacted to prevent intentional intrusion onto someone else's computer—specifically, computer hacking.'") (quoting *hiQ Labs, Inc. v. LinkedIn Corp.*, 31 F. 4th 1180, 1196 (9th Cir. 2022); *see also Healthcare Advocs., Inc. v. Harding, Earley, Follmer & Frailey*, 497 F. Supp. 2d 627, 642 (E.D. Pa. 2007) ("The DMCA makes it a violation of copyright law for a person to engage in activity commonly referred to as hacking when the object of that activity is to access copyrighted material that is protected by technological means.").

Because the scraping at issue only concerns public posts, none of the anti-hacking statutes X invokes apply.

B. X's CFAA and CDAFA Claims Fail Because Accessing a Public Website Is Not "Without Authorization."

The CFAA prevents intentional access to "a computer without authorization." 18 U.S.C. § 1030(a)(2)(C). X claims unauthorized access in three ways – violation of X's Terms, avoidance of X's "anti-scraping measures," and use of Bright Data's proxy IP network – but none rise to the level of "without authorization" under the CFAA. SAC ¶ 205. The Supreme Court counsels against reading the CFAA to "criminalize[] every violation of a computer-use policy." *Van Buren v. United States*, 593 U.S. 374, 394 (2021). Following *Van Buren*, the Ninth Circuit holds that conduct subject to the CFAA must be analogous to "breaking and entering." *hiQ*, 31 F.4th at 1197. X alleges nothing of the sort here. *See Ryanair DAC v. Booking Holdings Inc.*, 2024 WL 3732498, *6 (D. Del. 2024) ("The Ninth Circuit concluded that a business cannot transform a public website into a private one for purposes of the CFAA by implementing a ban on some users based on their perceived use of the website for commercial gain.").

Terms ≠ "Without Authorization." First, X says Bright Data accessed X's servers "without authorization and in violation of X Corp's Terms…." SAC ¶ 202. But, because "the CFAA is best understood as an anti-intrusion statute and not a misappropriation statute," the Ninth Circuit has long "rejected the contract-based interpretation of the CFAA…." *hiQ*, 31 F.4th at 1196-97 (collecting cases); *see also, e.g.*, *Chegg, Inc. v. Doe*, 2023 WL 4315540, *2 (N.D. Cal. 2023) ("[A] violation of the terms of use of a website—without more—cannot establish liability under the CFAA.").

X's Anti-Scraping Technologies ≠ "Without Authorization." X alleges that it employs "technological barriers" to prevent unwanted scraping of public information: "CAPTCHAs, login requirements, rate limits, and robots.txt restrictions." SAC ¶ 205.[2] None of these technologies turn public data search into hacking or unauthorized access to public websites.

As an initial matter, X alleges only that it uses CAPTCHAs as "account-creation restrictions," not restrictions on access to public information. SAC ¶¶ 27, 70, 79 ("X imposes a 'CAPTCHA' process to ensure that a human (rather than an automated process) is creating the account…. X Corp. requires potential registrants to pass a CAPTCHA by inputting the required information…"). Thus, a scraper can visit any public portion of X's site without ever encountering a CAPTCHA. Even if this could be an access restriction under the CFAA (which it is not), X does not allege Bright Data circumvented it.

As to login requirements, X does not plausibly allege any logged-in scraping by Bright Data. Its only logged-in scraping allegations are purely conclusory. SAC ¶¶ 122, 179, 205, 208.

[2] X also generally references "anomaly detection tools," but does not identify what they are or what Bright Data allegedly does to avoid them. SAC ¶ 205. In any event, the answer is the same for purposes of the CFAA: tools designed to detect certain kinds of public access do not make the access "without authorization."

Indeed, X acknowledges that Bright Data advertises "its datasets sourced from X contain only publicly accessible data," but alleges only disbelief in response, not actual facts that would render Bright Data's statements untrue. SAC ¶ 120. In short, X does not allege a single instance of scraping while logged in to any Bright Data or any other account, nor any piece of information collected by Bright Data that required a login.

Nor do rate limits make public data non-public. They simply "limit the amount of data that can be obtained over a set amount of time." *See* SAC ¶¶ 46, 67. The same with robots.txt. In *hiQ*, the Ninth Circuit rejected the argument that robots.txt and other technological anti-scraping measures can turn a public website non-public under the CFAA. *hiQ*, 31 F.4th at 1197 (These measures are not "authentication requirement[s], such as a password gate, [that] is needed to create the necessary barrier that divides open spaces from closed spaces on the Web.").

Proxy IPs ≠ "Without Authorization." Finally, X tries to repurpose its previously dismissed fraud-based UCL under the CFAA. SAC ¶¶ 207-209. But as this Court recognized, "there is no affirmative duty that requires an internet user to identify oneself with a given IP address when connecting to the internet." ECF 83 at 13. That ruling applies with equal force to X's CFAA claim.

X's CDAFA Claim Fails for the Same Reasons. Because "the necessary elements of" a CDAFA claim "do not differ materially from the necessary elements of the CFAA," "the CDAFA claim fails for the same reasons that the CFAA claim does." *Nowak v. Xapo, Inc.*, 2020 WL 6822888, *5 (N.D. Cal. 2020) (excepting the CFAA's requirement of $5,000 in damages); *see also Meta Platforms, Inc. v. BrandTotal Ltd.*, 605 F. Supp. 3d 1218, 1270 (N.D. Cal. 2022) (both the CFAA and CDAFA are "focused on hacking and not applicable to public websites").[3]

C. X's DMCA Claim Fails Because X's Technological Measures Are Not "Digital Walls" Controlling Access to Its Public Website.

The DMCA was, among other things, intended to give copyright holders tools against hacking and piracy in the digital world. *Microsoft Corp. v. AT & T Corp.*, 550 U.S. 437, 458 (2007) ("Congress addressed the ease with which pirates could copy and distribute a copyrightable work in digital form.") But Bright Data is neither pirate nor hacker. When accessing public information on X, Bright Data "access[es] the X platform as a normal user would." SAC ¶ 84.

X's Copyrighted Content Is Limited to Arrangement of Its Website. Section 1201(a)(1)(A) of the DMCA prohibits circumventing "a technological measure that effectively controls access to a work protected under this title." 17 U.S.C. § 1201(a)(1)(A). This "can be characterized as breaking and entering (or hacking) into computer systems." *Joint Stock Co. Channel One Russ. Worldwide v. Infomir LLC*, 2017 WL 696126, *18 (S.D.N.Y. 2017) ("Violating

[3] X cannot use the CDAFA to criminalize Bright Data's "use" of data. Where information is publicly accessible, X's Terms of Use cannot create CDAFA liability for only certain methods of use. *Oracle USA, Inc. v. Rimini Street, Inc.*, 879 F.3d 948, 962 (9th Cir. 2018), *rev'd on other grounds*, 586 U.S. 334 (2019) ("Taking data using a method prohibited by the applicable terms of use, when the taking itself generally is permitted, does not violate the CDAFA.").

contractual restrictions on access to or distribution of encrypted transmissions is not the type of 'circumvention' that Congress intended to combat in passing the DMCA."). This, X does not allege.

X does not assert technological measures to protect public user content on X (and indeed disclaims any copyright in such content), but instead asserts copyright in X's "websites, including twitter.com and X.com, and mobile and online applications." SAC ¶ 190. X's DMCA claim is thus premised on its alleged copyright in the "selection, coordination, or arrangement of the content appearing on" its website, or perhaps "the overall hierarchy of the website."[4] X admits its website is publicly available to, for example, "users who navigate to X's website from another source, such as a search engine or an X post that is embedded within another website." SAC ¶ 64. Such logged-out members of the public can then see a poster's profile, "a curated list of other posts from that individual," and "other linked posts." SAC ¶¶ 66-67. As noted, X has not plausibly alleged that Bright Data accessed anything other than this publicly available content.

Instead, X claims its copyrighted content is protected by "CAPTCHA, login requirements, rate limits, robots.txt restrictions, and anomaly detection tools." SAC ¶ 191. But none control access to X's copyrighted content (the selection, coordination, or arrangement of its website), because that information is publicly available.

Bright Data Did Not Circumvent CAPTCHAs. As noted, X's CAPTCHAs do not control access to its website, but are used only as account verification tools. X does not allege that a logged-out internet surfer accessing X's public webpage is ever shown a CAPTCHA. Thus, X does not allege that Bright Data solved a CAPTCHA (though solving would not be circumvention), or was even prompted to solve a CAPTCHA to access X's public website. It is not a technological measure to protect X's asserted copyrighted work, and it was not circumvented.

Bright Data Did Not Circumvent Login Requirements. As to "login requirements," X does not allege that Bright Data entered an unauthorized or fraudulent password to hack into X's site. *See Aeropost Int'l Servs. v. Aerocasillas, S.A.*, 2011 WL 13174672, *6 (S.D. Fla. 2011) (dismissing DMCA claim where "the allegations are that Defendants used a valid login and password to enter the website and to create another login, which they were freely able to do without any technological impediment"); *iSpot.tv, Inc. v. Teyfukova*, 2023 WL 1967958, *12 (C.D. Cal. 2023) ("The allegations in the FAC indicate that, rather than avoiding or bypassing the system, Teyfukova simply used the Horizon credentials as they were intended to be used."); *see also Schork Grp., Inc. v. Choice! Energy Servs., Retail, LP*, 2022 WL 2905231, *10 (E.D. Pa. 2022) (even "misuse of a legitimate password does not fall within the ambit of" the DMCA).

Rather, the DMCA is designed for situations where "a defendant uses some technology to open the digital lock, other than the password specifically made for the digital lock, [thereby] avoiding or bypassing the digital lock." *Burroughs Payment Sys. v. Symco Grp.*, 2011 WL 13217738, *5 (N.D. Ga. 2011) ("To find otherwise would be to read 'avoid' and 'bypass' as

[4] U.S. Copyright Office, Circular 66, *Copyright Registration of Websites and Website Content*, at 3 https://www.copyright.gov/circs/circ66.pdf.

including the concept of unauthorized 'use.' This reading is not supported by the statute, nor is it supported by the common meanings of these words."). X does not identify any such avoidance or bypass.

Rate Limits Are Neither "Effective Controls" Nor Circumvented. Nor are X's rate limiters effective controls under the DMCA. A "technological measure" under the DMCA is one that "in the ordinary course of its operation, requires the application of information, or a process or a treatment, … to gain access to the work." 17 U.S.C. § 1201(a)(3)(B). Rate limiters do not prevent access to the website; just the frequency with which such information can be accessed. They work by first "identifying the IP [address] … of the individual requesting access." SAC ¶ 192. Then, after some unspecified number of clicks on public information, the rate limiter triggers a log-in or account creation prompt. SAC ¶ 67. But the allegedly copyrighted content has already been accessed. X says Bright Data circumvents this, not by applying some unauthorized information (such as hacking into an account), but simply by switching IP addresses and restarting the rate limiter. SAC ¶ 120. This is no different than continuing the search on a different device. Bright Data doesn't break a lock, it uses an open path. *See Couponcabin LLC v. Savings.com, Inc.*, 2016 WL 3181826, *6 (N.D. Ind. 2016) (DMCA claim dismissed where plaintiff "alleges that the Defendants continued their scraping activity by accessing the Plaintiff's website using a variety of servers and/or internet service providers" because plaintiff's "technological safeguards and barriers" thus did not amount to "effective[] controls" under the DMCA).

Robots.txt Are Not "Effective Controls." Robots.txt fares no better. A few months before X filed suit, X "modified its robots.txt instructions to prohibit all forms of automated access to X's website except for Google's web crawler." SAC ¶ 75. But "no court has found that a robots.txt file universally constitutes a 'technological measure effectively controll[ing] access' under the DMCA." *Healthcare Advocs., Inc.*, 497 F. Supp. 2d at 643. This is because "[a]dherence to the rules in a robots.txt file is voluntary." *hiQ Labs, Inc. v. LinkedIn Corp.*, 938 F.3d 985, 992 n.2 (9th Cir. 2019), *cert. granted, judgment vacated*, 141 S. Ct. 2752 (2021). A voluntary measure is simply not the digital lock the DMCA was intended to protect. *Healthcare Advocs., Inc.*, 497 F. Supp. 2d at 645-46.

Sincerely,

/s *Colin Kass*